Haozhu Wang

Project 3.1 Report

In this assignment we are to do association analysis on the weather data. I must run all the association learner algorithm of WEKA on it. They are Apriori, PredictiveApriori, Tertius, and HotSpot.

Apriori- algorithm reduces the minimum support until it finds the required number of rules with given minimum confidence. Then it mines class association rules.

Predictive Apriori - It searches with an increasing support threshold for the best 'n' rules concerning a support-based corrected confidence value. It searches with an increasing support threshold for the best.

Tertius- Finds rules according to confirmation measure.

HotSpot- This learner learns a set of rules (displayed in a tree-like structure) that maximize/minimize a target variable/value of interest. With a nominal target, one might want to look for segments of the data where there is a high probability of a minority value occuring (given the constraint of a minimum support). For a numeric target, one might be interested in finding segments where this is higher on average than in the whole data set.

Results:

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -V -c -1
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===

Apriori
=======
Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 conf:(1)
2. temperature=cool 4 ==> humidity=normal 4 conf:(1)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 conf:(1)
4. outlook=sunny play=no 3 ==> humidity=high 3 conf:(1)
5. outlook=sunny humidity=high 3 ==> play=no 3 conf:(1)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3 conf:(1)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 conf:(1)
8. temperature=cool play=yes 3 ==> humidity=normal 3 conf:(1)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2 conf:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2 conf:(1)

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 25 -c -1
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===
PredictiveApriori
==========

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 acc:(0.95323)
2. temperature=cool 4 ==> humidity=normal 4 acc:(0.95323)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 acc:(0.95323)
4. outlook=sunny humidity=high 3 ==> play=no 3 acc:(0.92093)
5. outlook=sunny play=no 3 ==> humidity=high 3 acc:(0.92093)
6. outlook=rainy windy=FALSE 3 ==> play=yes 3 acc:(0.92093)
7. outlook=rainy play=yes 3 ==> windy=FALSE 3 acc:(0.92093)

8. outlook=sunny temperature=hot 2 ==> humidity=high 2 acc:(0.86233)
9. outlook=sunny humidity=normal 2 ==> play=yes 2 acc:(0.86233)
10. outlook=sunny play=yes 2 ==> humidity=normal 2 acc:(0.86233)
11. outlook=overcast temperature=hot 2 ==> windy=FALSE 2 acc:(0.86233)
12. outlook=overcast windy=FALSE 2 ==> temperature=hot 2 acc:(0.86233)
13. outlook=rainy play=no 2 ==> windy=TRUE 2 acc:(0.86233)
14. humidity=normal 7 ==> play=yes 6 acc:(0.69497)
15. play=no 5 ==> humidity=high 4 acc:(0.59096)
16. windy=FALSE 8 ==> play=yes 6 acc:(0.56435)
17. temperature=hot 4 ==> humidity=high 3 acc:(0.54473)
18. temperature=hot 4 ==> windy=FALSE 3 acc:(0.54473)
19. temperature=cool 4 ==> humidity=normal play=yes 3 acc:(0.54473)
20. play=yes 9 ==> humidity=normal 6 acc:(0.53808)
21. play=yes 9 ==> windy=FALSE 6 acc:(0.53808)
22. temperature=mild 6 ==> humidity=high 4 acc:(0.51949)
23. temperature=mild 6 ==> play=yes 4 acc:(0.51949)
24. outlook=sunny humidity=high 3 ==> temperature=hot 2 acc:(0.49529)
25. windy=TRUE play=yes 3 ==> temperature=mild 2 acc:(0.49529)

=== Run information ===

Scheme: weka.associations.Tertius -K 10 -F 0.0 -N 1.0 -L 4 -G 0 -c 0 -I 0 -P 0
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===

Tertius
=======

1. /* 0.633754 0.071429 */ play = yes ==> humidity = normal or outlook = overcast
2. /* 0.607625 0.000000 */ humidity = normal ==> temperature = cool or play = yes
3. /* 0.607625 0.000000 */ temperature = cool ==> humidity = normal
4. /* 0.594071 0.214286 */ humidity = normal ==> temperature = cool
5. /* 0.590214 0.000000 */ humidity = high and outlook = sunny ==> play = no
6. /* 0.555556 0.000000 */ play = no ==> windy = TRUE or outlook = sunny
7. /* 0.486606 0.000000 */ play = no and outlook = sunny ==> humidity = high
8. /* 0.486606 0.000000 */ humidity = normal ==> play = yes or outlook = rainy
9. /* 0.469374 0.000000 */ outlook = overcast ==> play = yes

10. /* 0.469374 0.000000 */ windy = FALSE and outlook = overcast ==> temperature = hot
11. /* 0.469374 0.000000 */ outlook = overcast ==> temperature = hot or windy = TRUE
12. /* 0.469374 0.000000 */ temperature = hot and play = yes ==> outlook = overcast
13. /* 0.469374 0.000000 */ play = no ==> humidity = high or windy = TRUE
14. /* 0.469374 0.000000 */ temperature = hot ==> play = no or outlook = overcast
15. /* 0.469374 0.000000 */ temperature = hot ==> humidity = high or outlook = overcast
16. /* 0.469374 0.000000 */ humidity = high and play = no ==> temperature = mild or outlook = sunny
17. /* 0.469374 0.000000 */ temperature = mild and play = yes ==> windy = TRUE or outlook = rainy
18. /* 0.469374 0.000000 */ outlook = sunny ==> temperature = cool or windy = TRUE or play = no
19. /* 0.467119 0.357143 */ play = yes ==> outlook = overcast
20. /* 0.458333 0.071429 */ play = yes ==> windy = FALSE or outlook = overcast
21. /* 0.458333 0.071429 */ humidity = high and play = no ==> outlook = sunny
22. /* 0.439100 0.071429 */ play = no ==> humidity = high
23. /* 0.439100 0.071429 */ humidity = high ==> temperature = mild or play = no
24. /* 0.439100 0.071429 */ humidity = high ==> temperature = mild or outlook = sunny

Number of hypotheses considered: 1724
Number of hypotheses explored: 689

=== Run information ===

Scheme: weka.associations.HotSpot -c last -V first -S 0.33 -M 2 -I 0.01
Relation: weather
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===

Hot Spot
========
Total population: 14 instances
Target attribute: play
Target value: yes [value count in total population: 9 instances (64.29%)]
Minimum value count for segments: 5 instances (33% of total population)
Maximum branching factor: 2
Minimum improvement in target: 1%

```
play=yes (64.29% [9/14])
  humidity <= 80 (85.71% [6/7])
  |   temperature > 65 (100% [5/5])
  windy = FALSE (75% [6/8])
  |   temperature <= 83 (85.71% [6/7])
  |   |   humidity <= 86 (100% [5/5])
  |   humidity <= 86 (83.33% [5/6])
```

It seems for most learners, with exception of HotSpot, recognizes the obvious relationship between outlook overcast with “yes” to play. HotSpot only did the learning on the numerical attributes. In Tertius learner, there are rules and associations being repeated probably due to similar hypothesis. However, due to the small data size, algorithms might draw associations that doesn’t make much logic sense in real life because of certain coincidences in the data. Whether you play golf or not should not have any influence on whether the temperature level or the humidity level, but the learners doesn’t differentiate difference contexts.